Exploratory Data Analysis on the Automobile Data Set

Introduction

This Report presents an analysis of an Automobile Dataset. The Dataset includes information on Fuel Type, city mpg, highway mpg, engine size, price, make, and body styles as indicated in the Data frame. The Purpose of this Analysis is to explore possible relationships and trends within the Dataset and Provide insights from the data, into the fuel efficiency of the cars in the dataset.

SUMMARY OF THE DATASET

The Dataset contains 201 rows and 7 columns after The Data Cleaning process. The Columns include both categorical and numerical data types. The Dataset has information on Fuel Type, city mpg, highway mpg, engine size, price, make, and body styles.

DATA CLEANING

During this process, firstly removed columns that will not be relevant to our analysis. Then checked for duplicates but we did not find any, so we went on to replace "?" with NaN, dropped rows with NaN then converted the price column to numeric data type.

MISSING DATA

When we checked if the are missing values, we find out that the are no missing values.

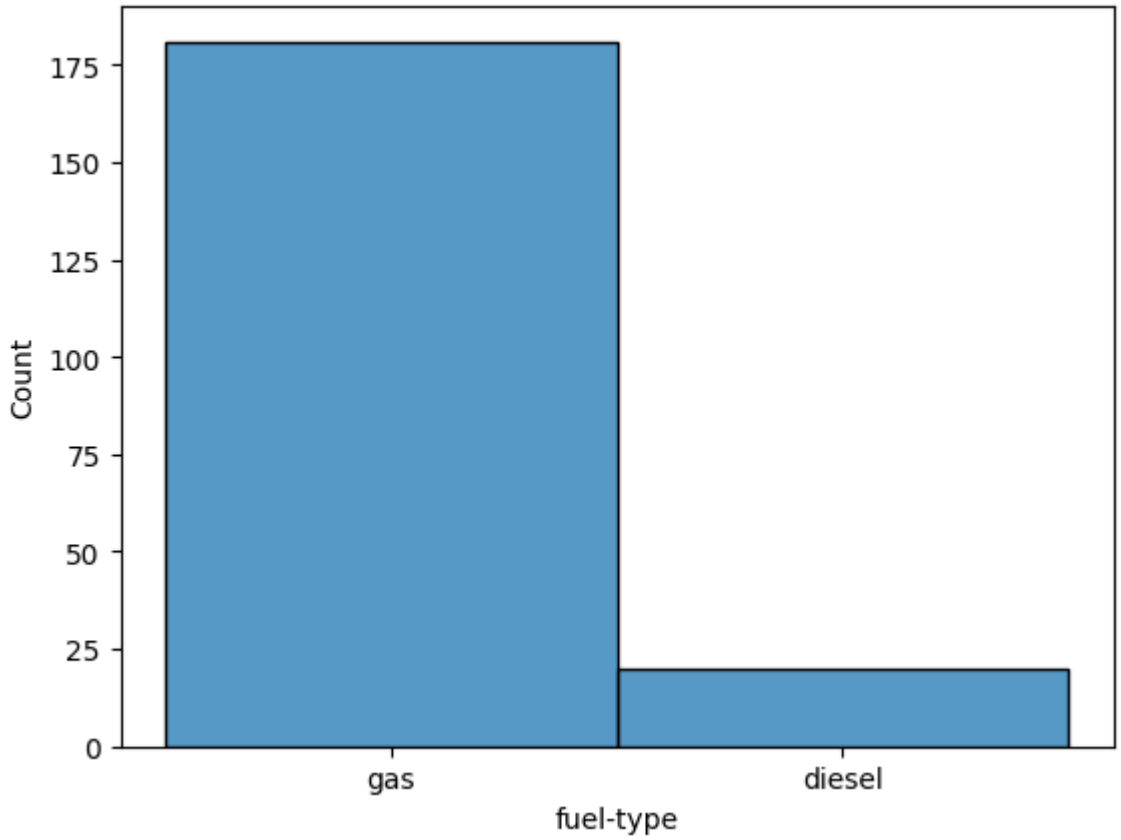
DATA STORIES AND VISUALISATIONS

We used various methods and visualisations to explore the dataset. We conducted univariate and bivariate analyses to understand the relationship between different variables. We used visualisations such as histograms, scatterplots, boxplots, count plots, bar plots, line plots and bar charts to help us better understand the data, and the following findings were made:

1. Univariate Analysis

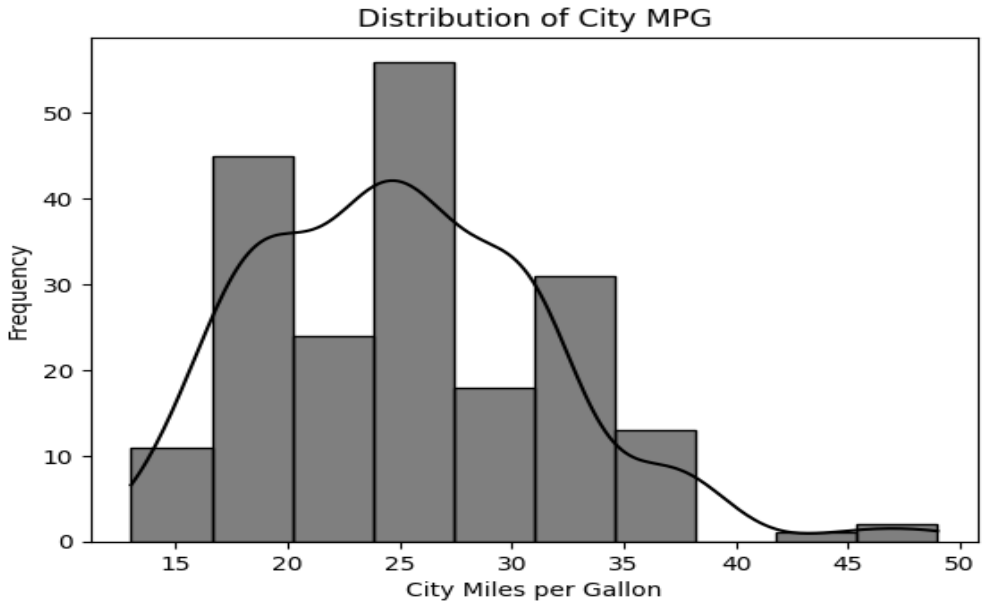
The Purpose of this analysis is to examine only one variable at a time.

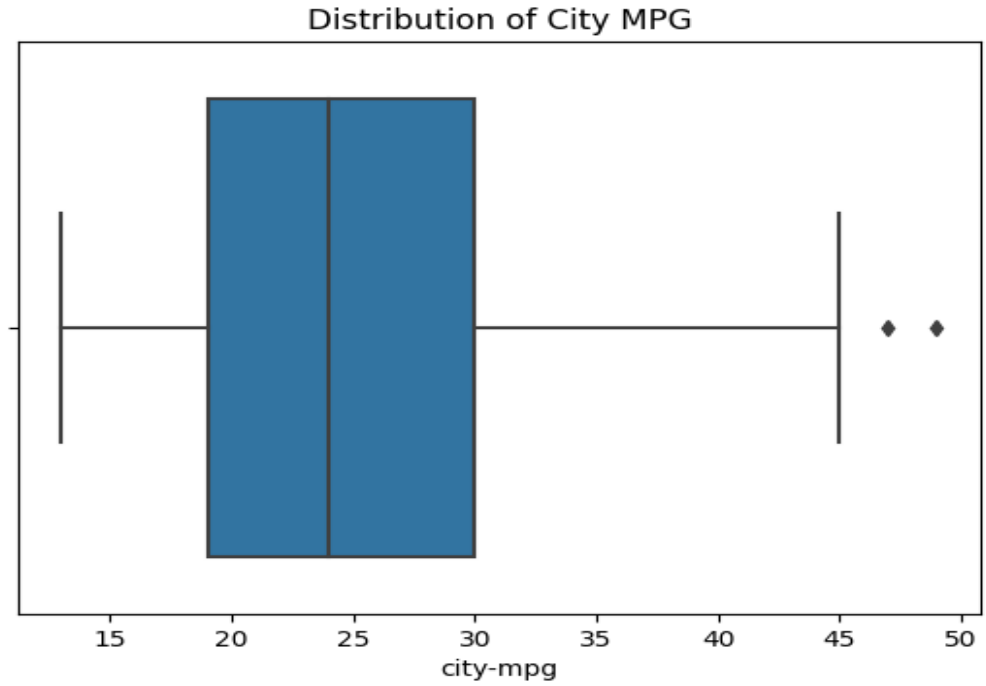
Distribution of Fuel Types (Histogram)



Majority of the cars in the Dataset use gas rather than diesel. The purpose was to assess which fuel type is dominant in the dataset and how that might affect our overall analysis of the cars in relation to fuel efficiency.

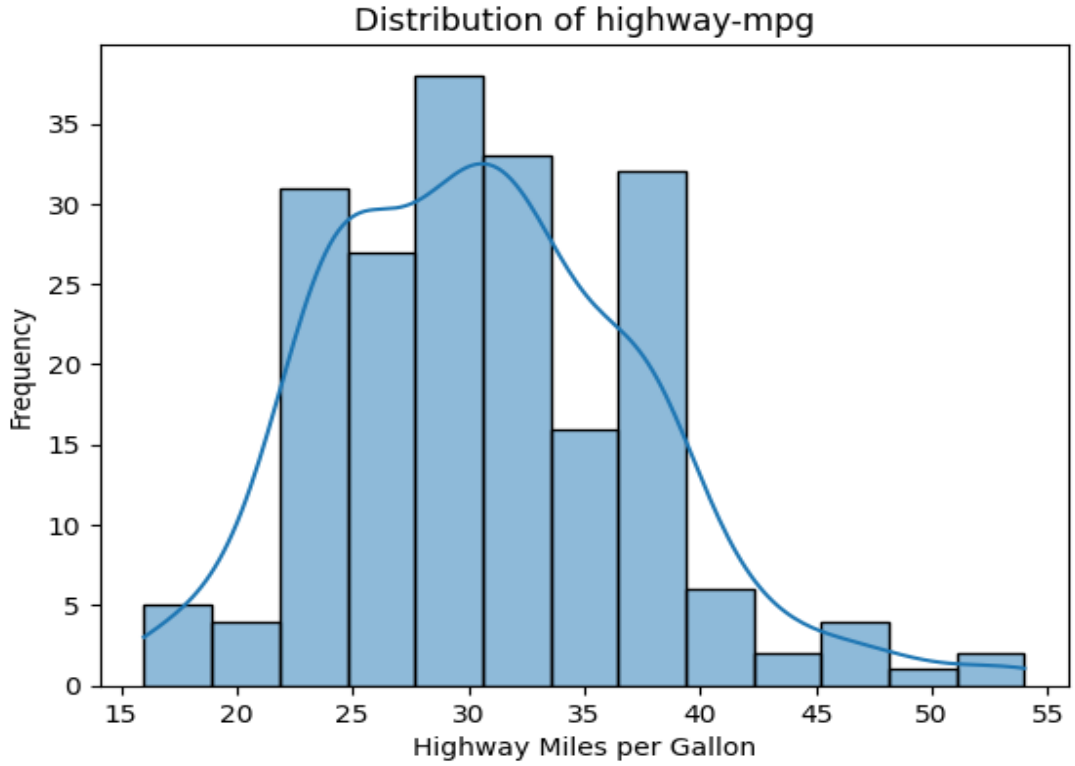
Distribution of City MPG (Histogram and Boxplot)

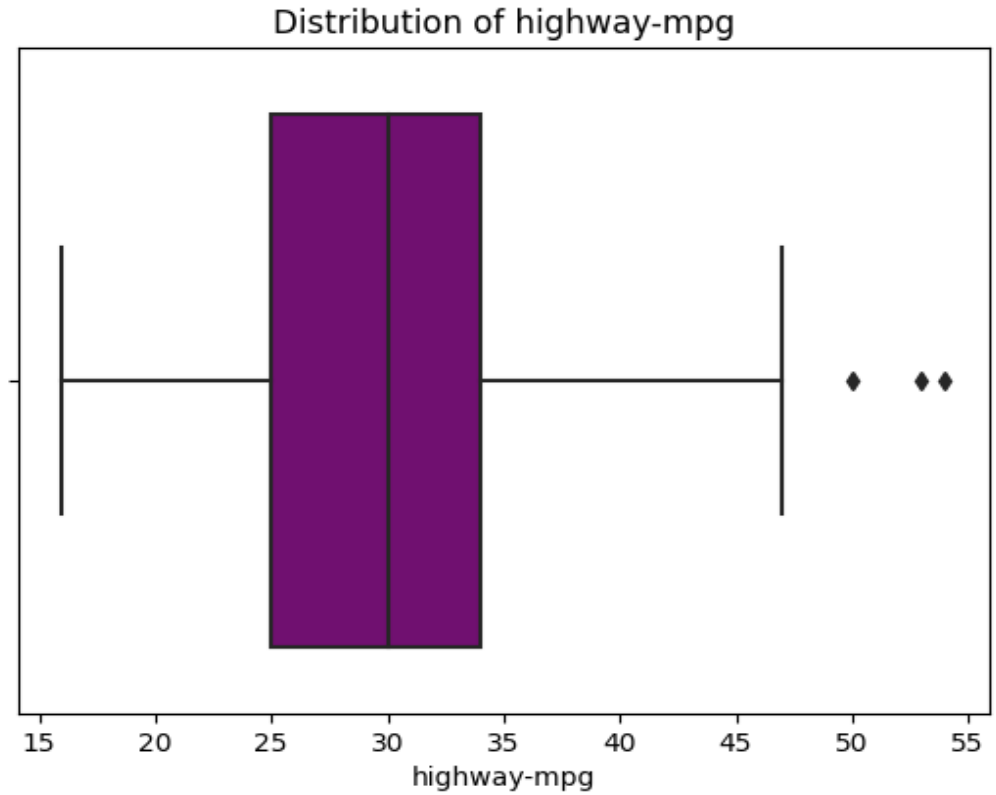




Majority of the cars are fuel efficient while driving in city traffic, this is supported by the positive skewed boxplot and the distribution line in the histogram. Most of them scored above average.

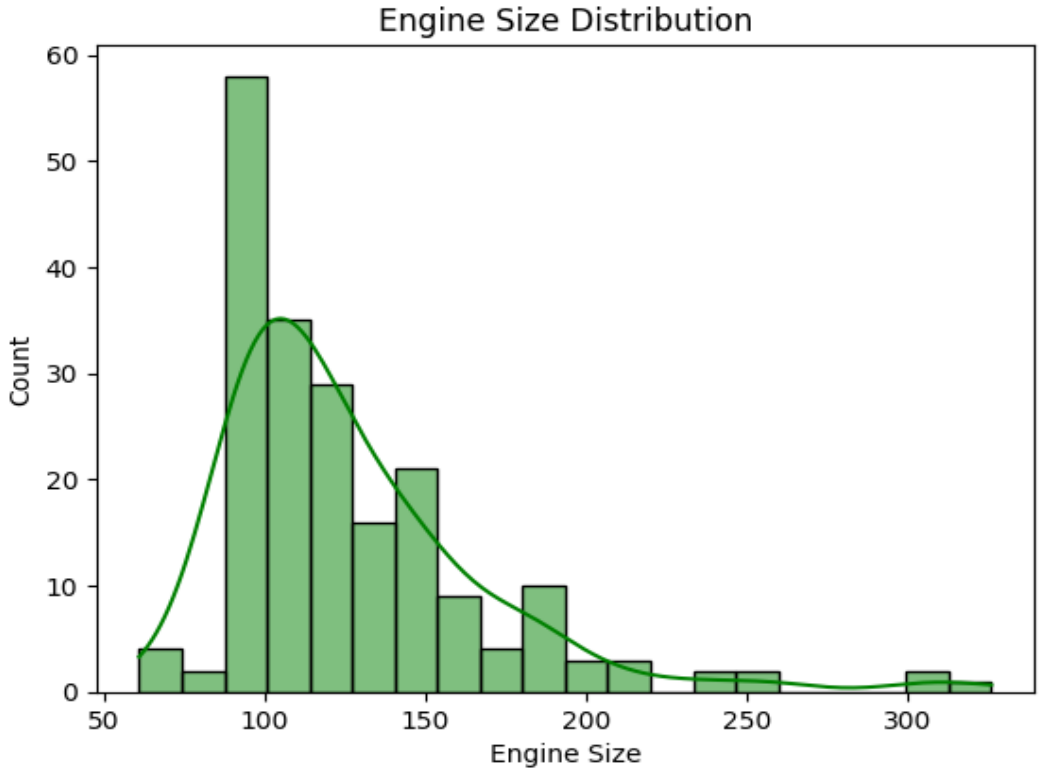
Distribution of Highway MPG (Histogram and Boxplot)

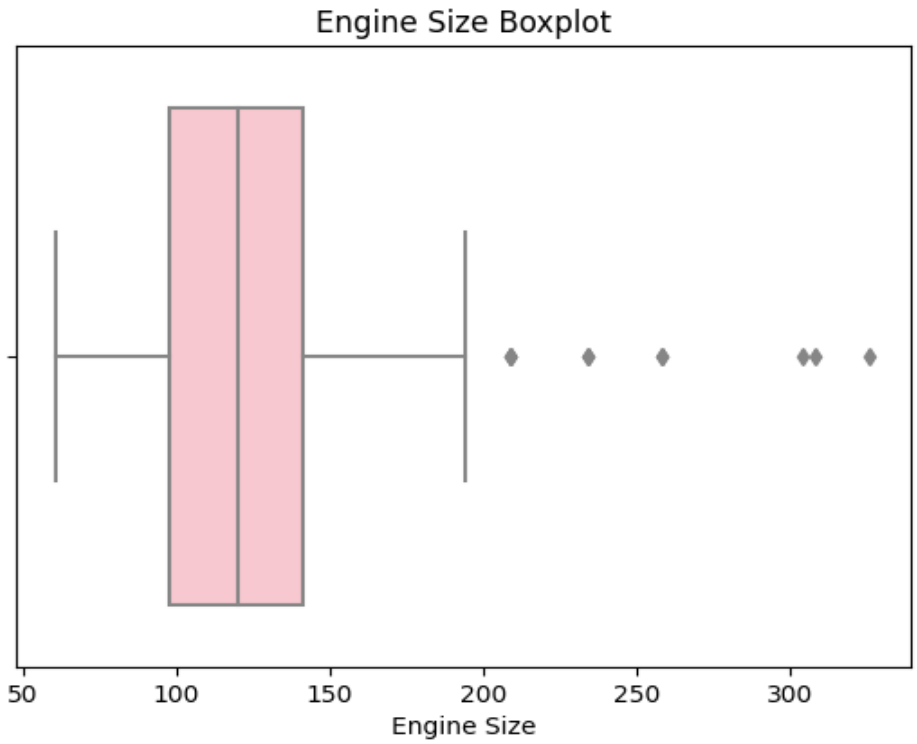




Contrary to the previous finding, most cars are not that fuel efficient while driving in highway, as indicated by the negatively skewed boxplot and the distribution line in the histogram. As most of them scored below average. However, the difference is not that huge from the City Mpg

Distribution of Engine size (Histogram and Boxplot)

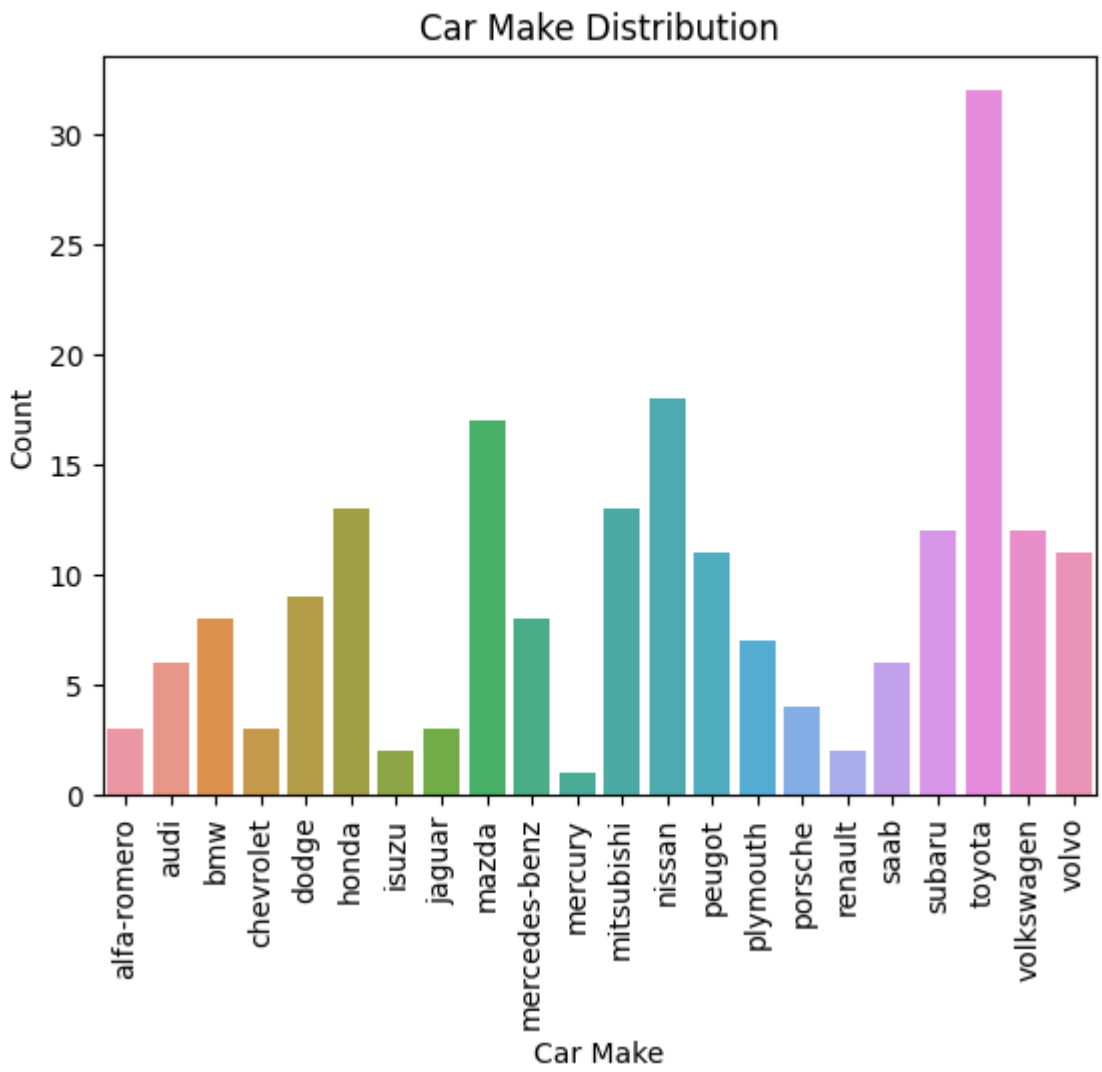




Majority of the cars in the dataset have small engines, cars with a large engine are present in the set however they are a minority. This is supported by negatively skewed boxplot and the distribution line in the histogram is at its highest in low values.

Car Make Distribution (Count plot):

.

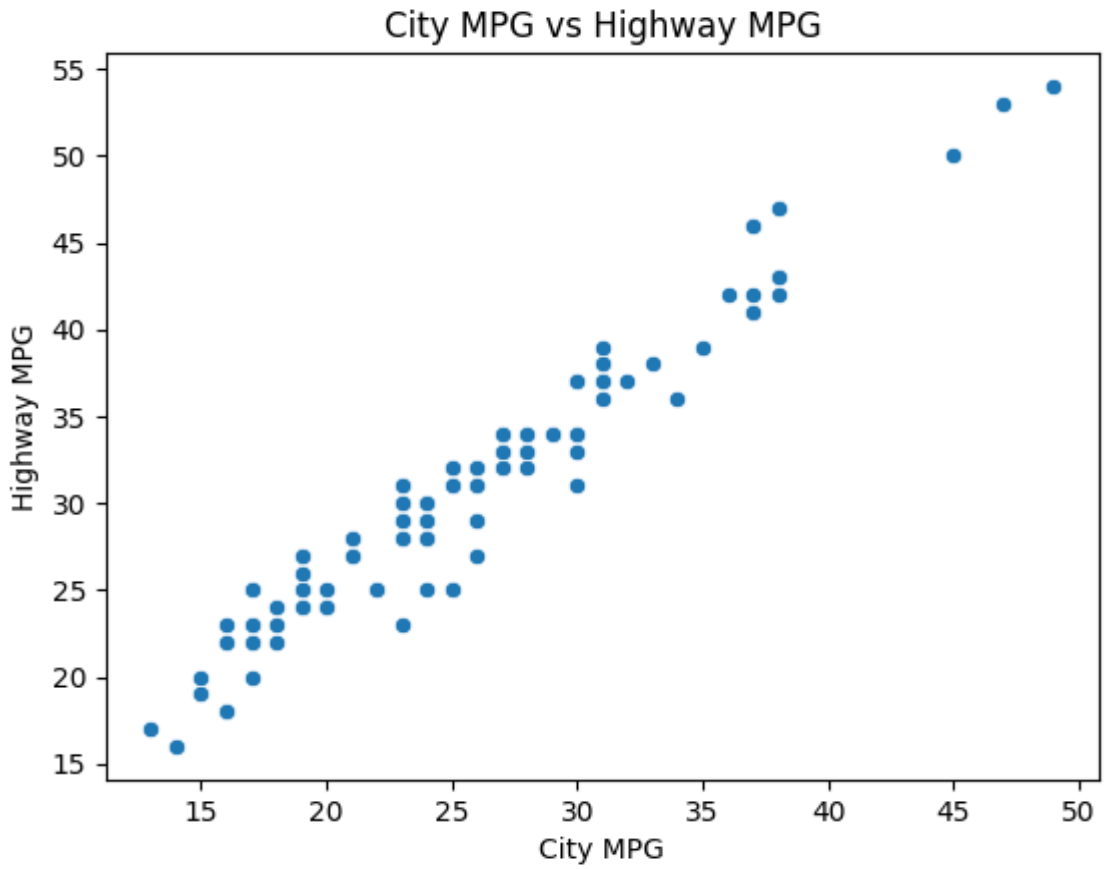


Toyota is the most Frequent Car make in the dataset, so most of the cars in the dataset are manufactured by Toyota. We also find out that most of the cars in the dataset are sedan body styles.

2. Bivariate Analysis

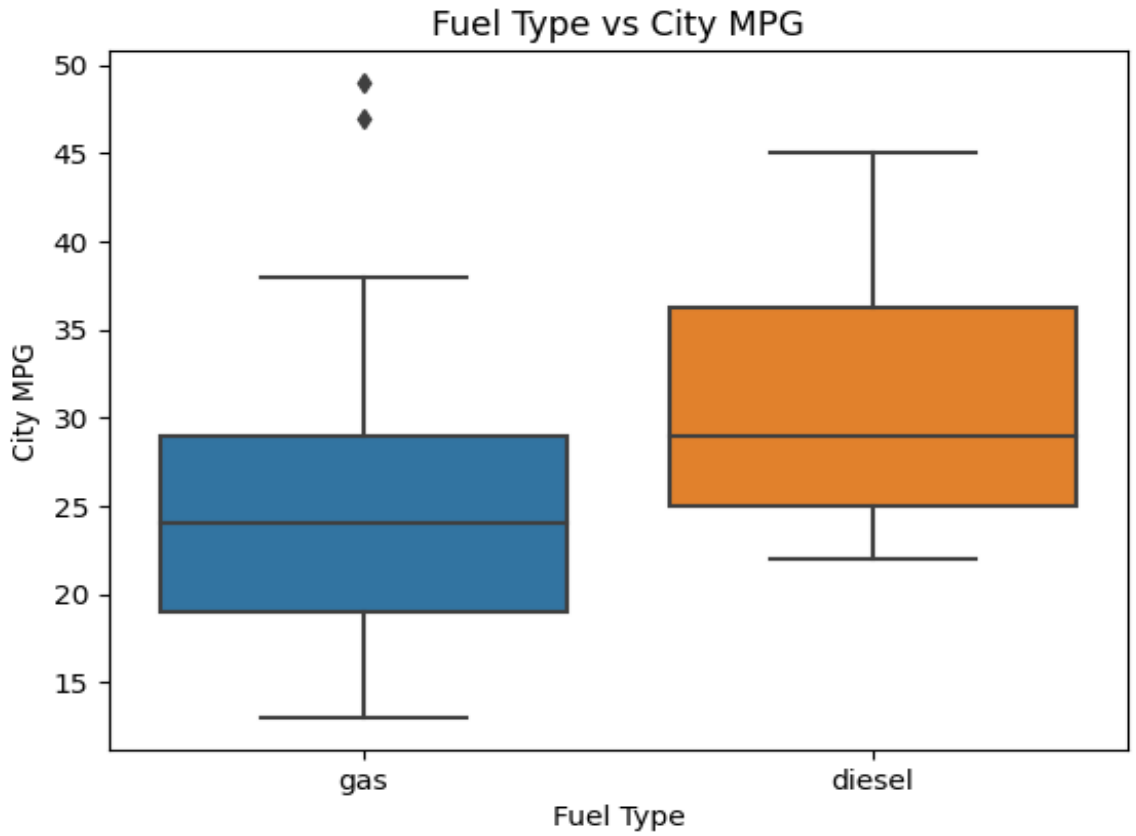
The purpose of analysis is to explore Relationships between variables in the Dataset.

City MPG vs highway MPG(Scatterplot)



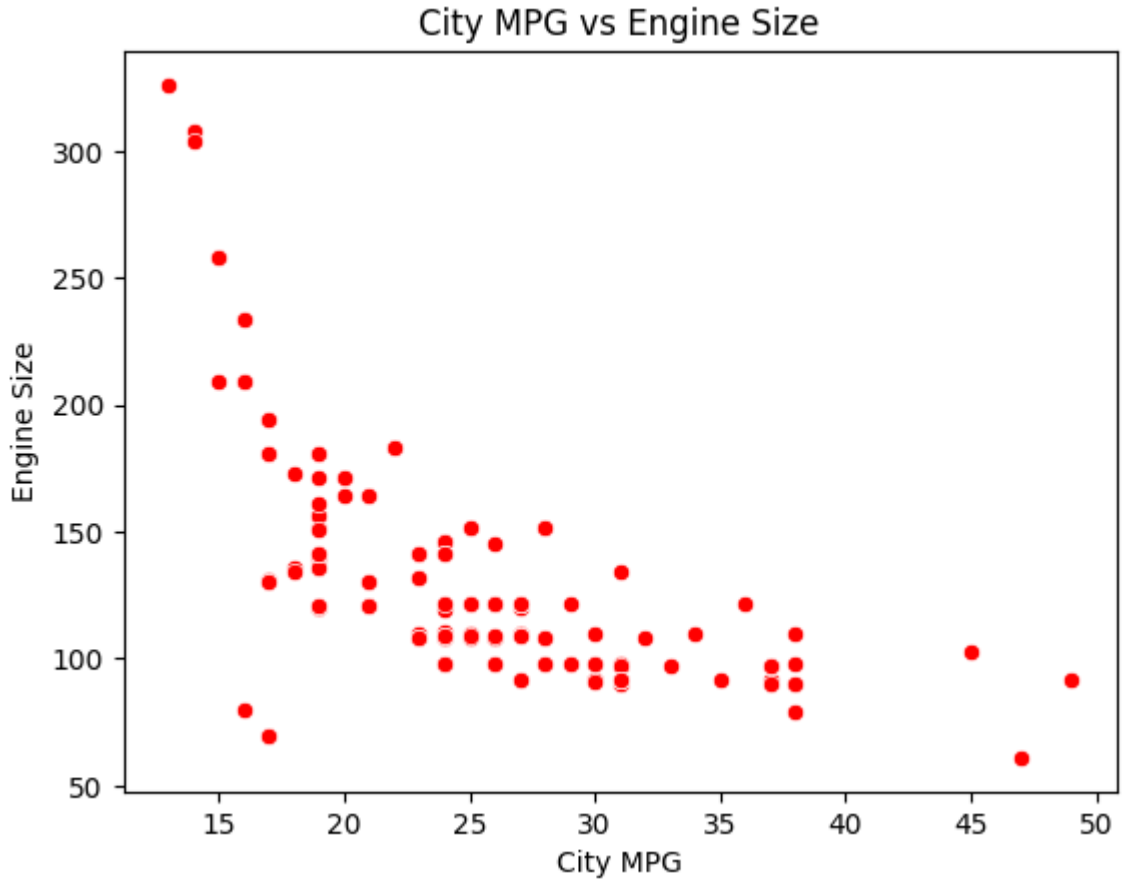
The is a positive correlation between the two variables, meaning they share a positive relationship. So, because of that we can only compare one MPG with the other variables, which at this case will be City Mpg. This is because the is straight uptrend indicated in the plot.

City MPG vs Fuel Type (Boxplot)



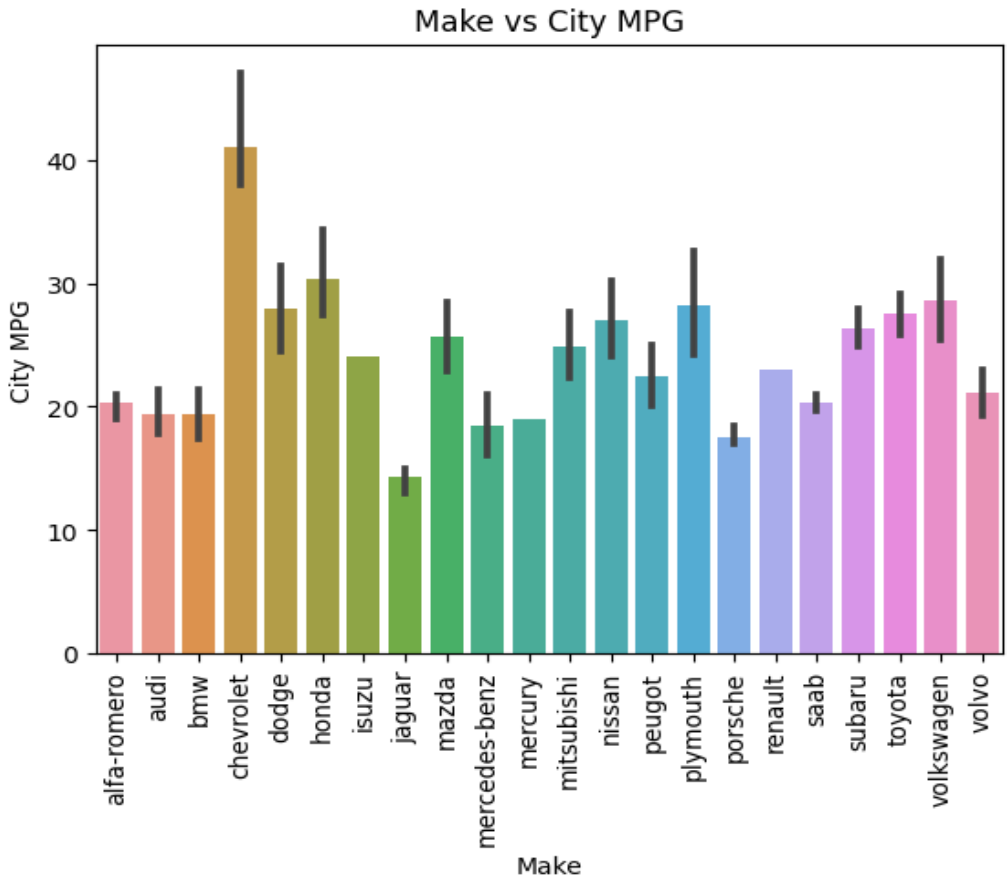
Cars that use diesel are more fuel efficient than cars that use gas. It is also worth noting that the plot supports the argument made that most of the cars in the dataset are fuel efficient.

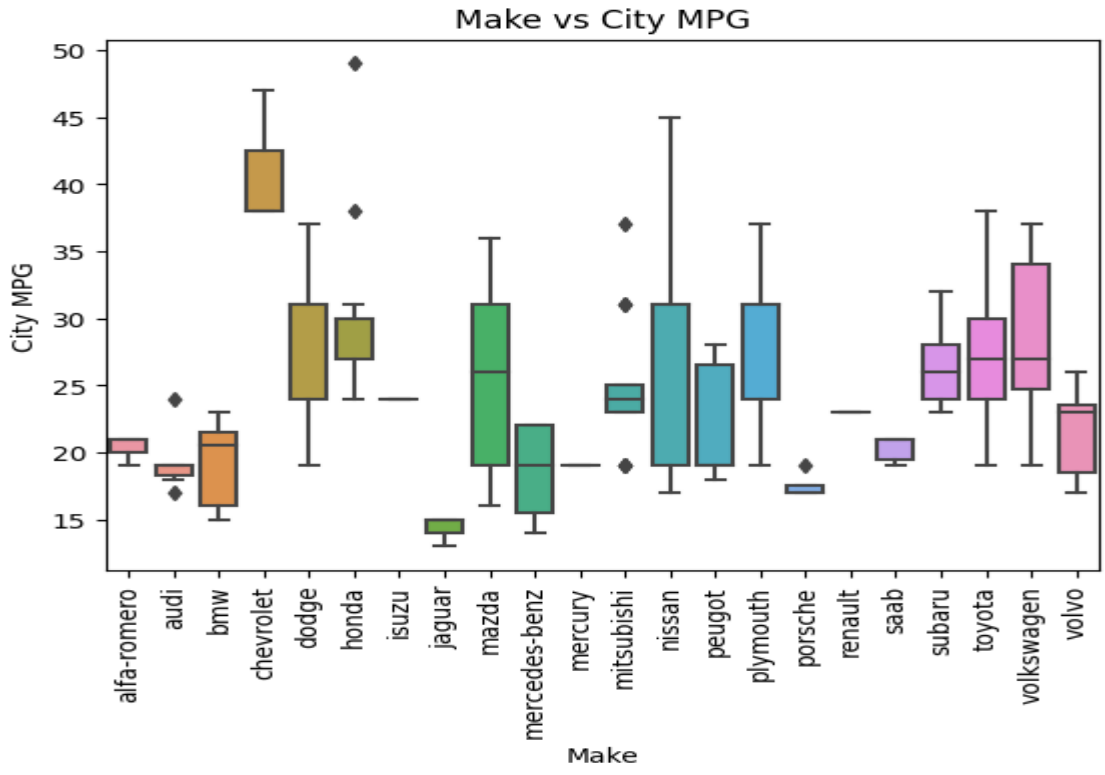
City MPG vs Engine Size (Scatterplot)

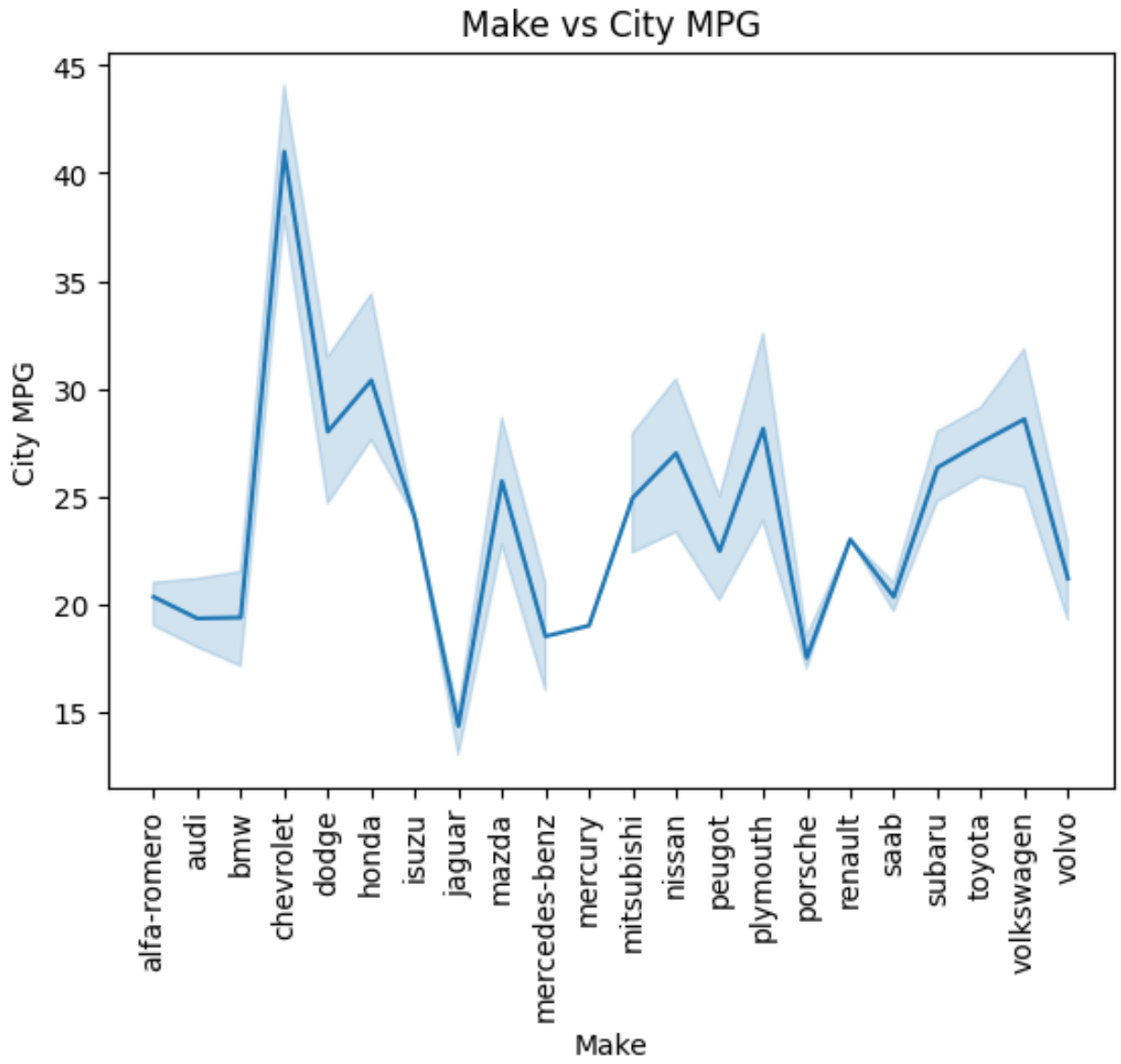


A strong negative relationship was discovered between the variables, which means that for every increase in engine size, the is a decline in fuel efficiency. So, the size of a car's engine does impact fuel efficiency. Cars that have big engine sizes are less fuel efficient. A correlation coefficient of - 0.65 supports this argument.

City MPG vs Make (Bar plot, Box plot and Line plot)

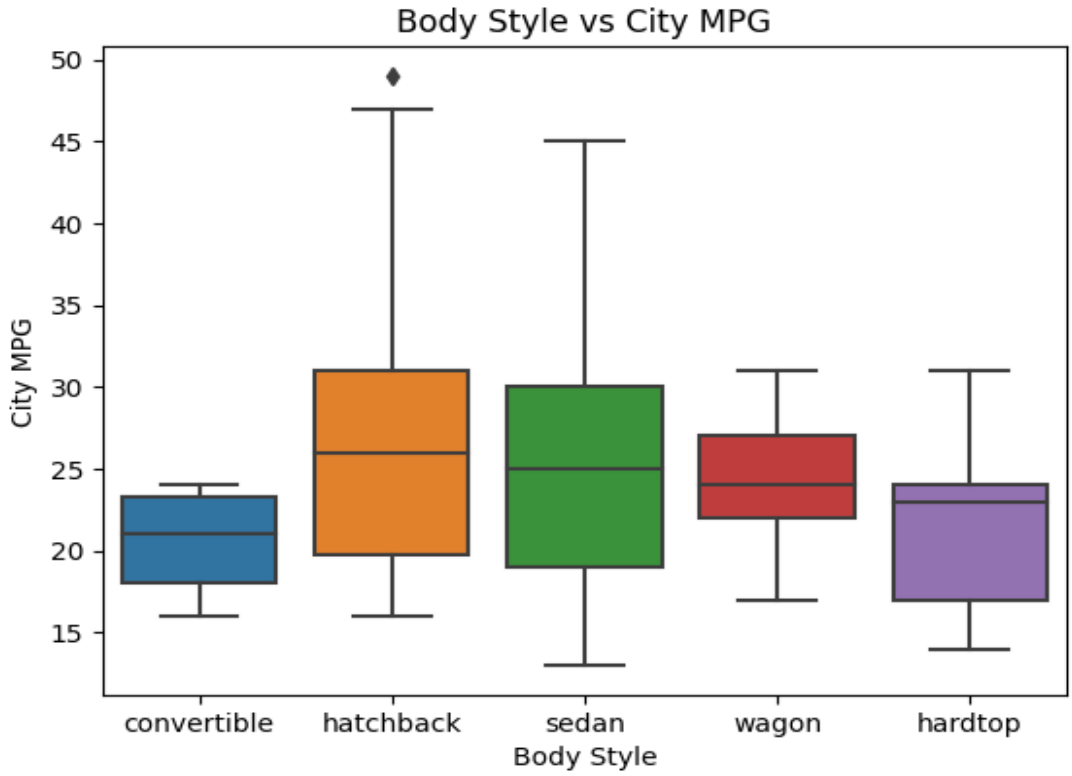


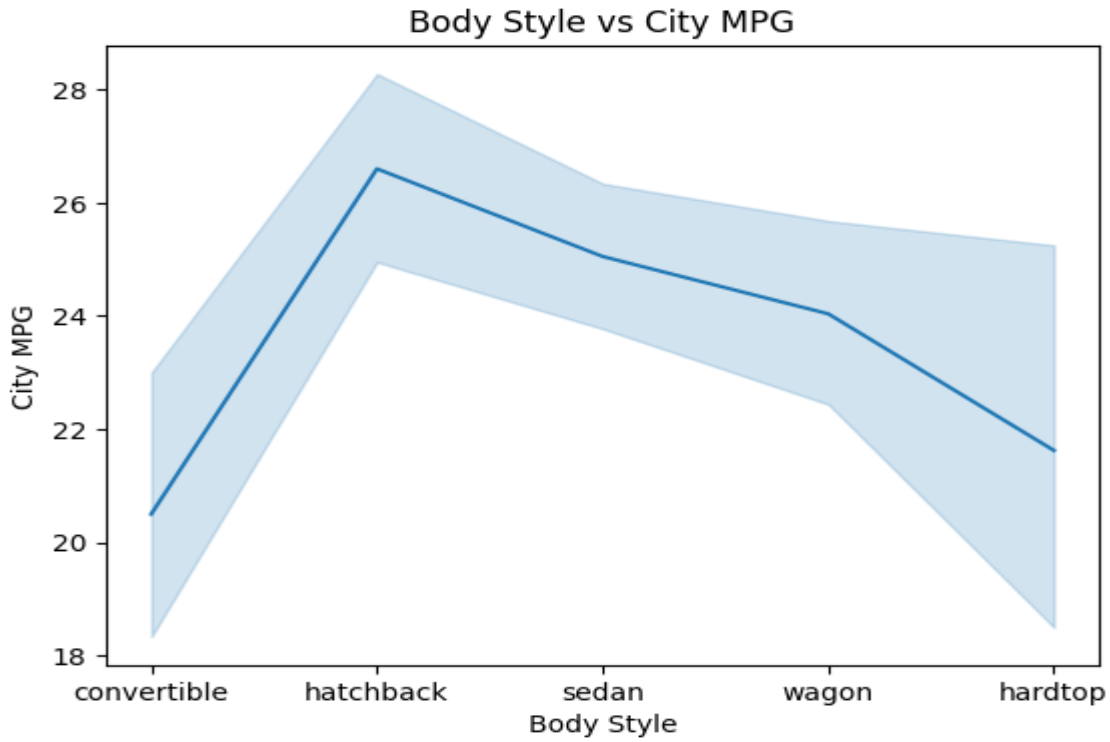


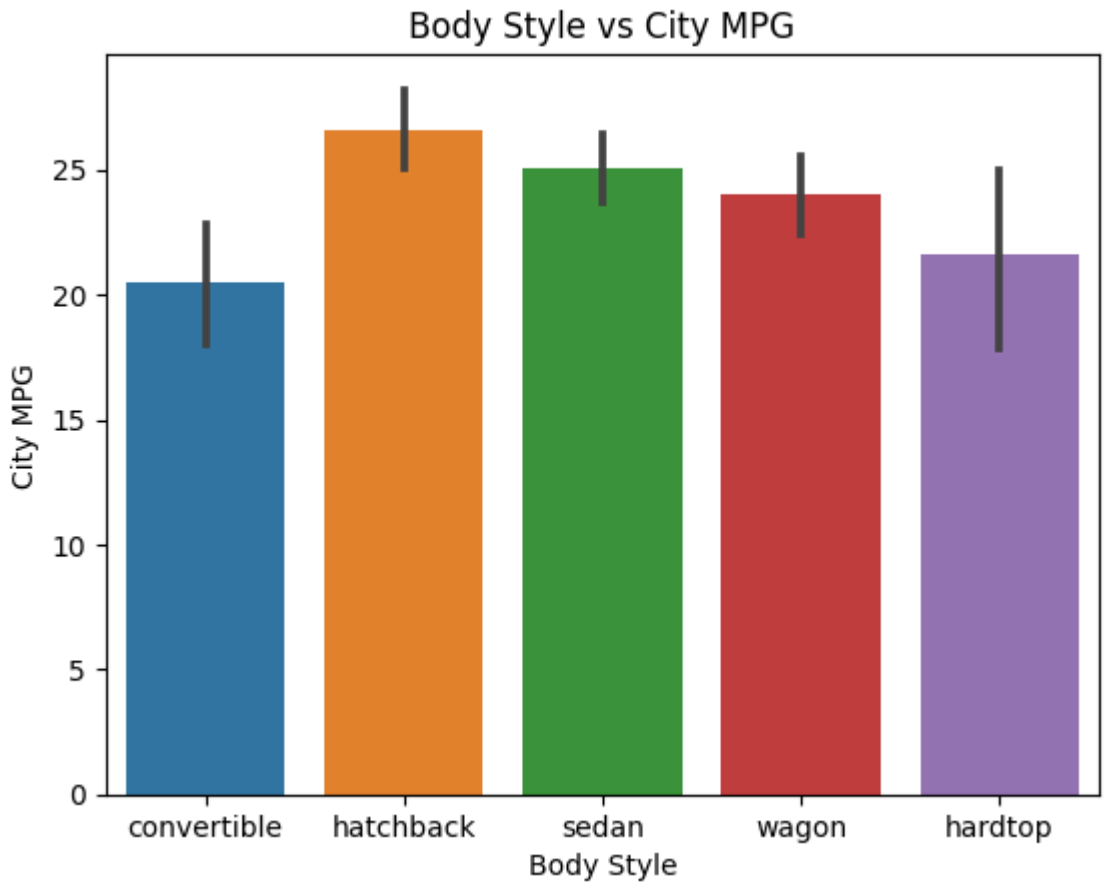


Chevrolet is the most fuel-efficient car make in the dataset, followed by Isuzu with Jaguar being the least fuel-efficient car in the Dataset.

City MPG vs Body Style (Bar plot, Box plot and Line plot):

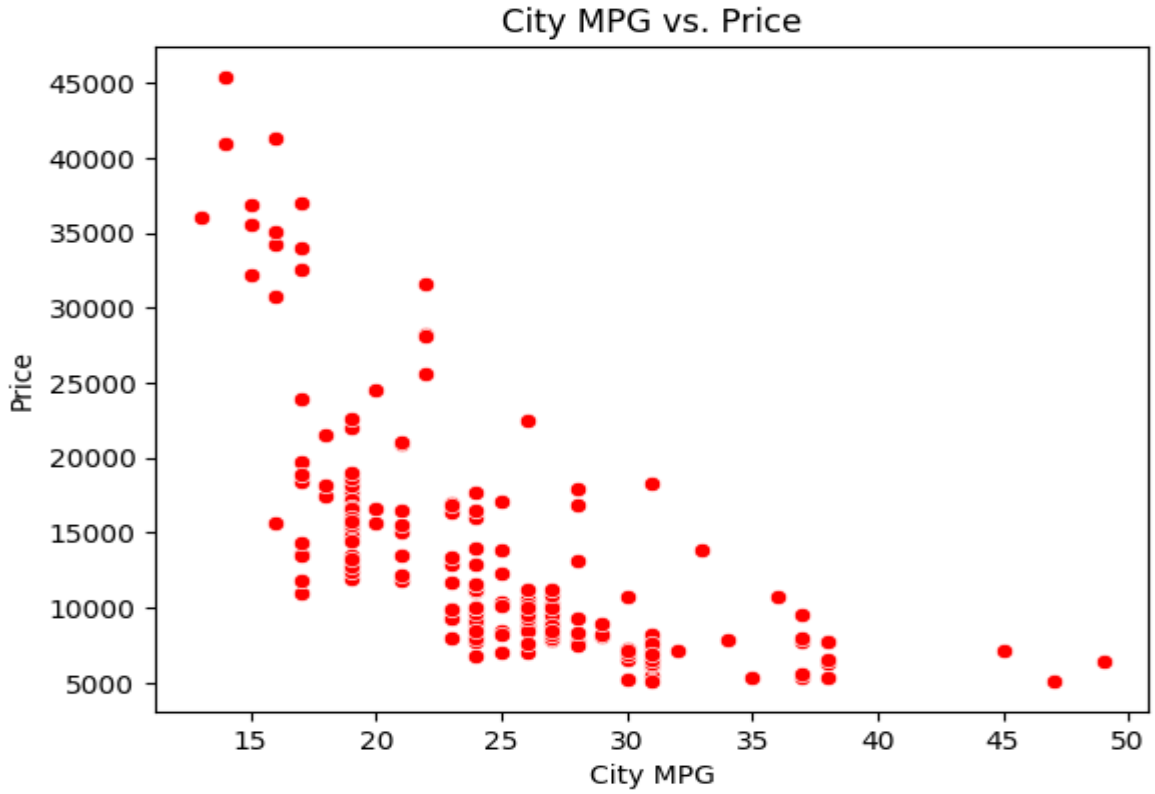


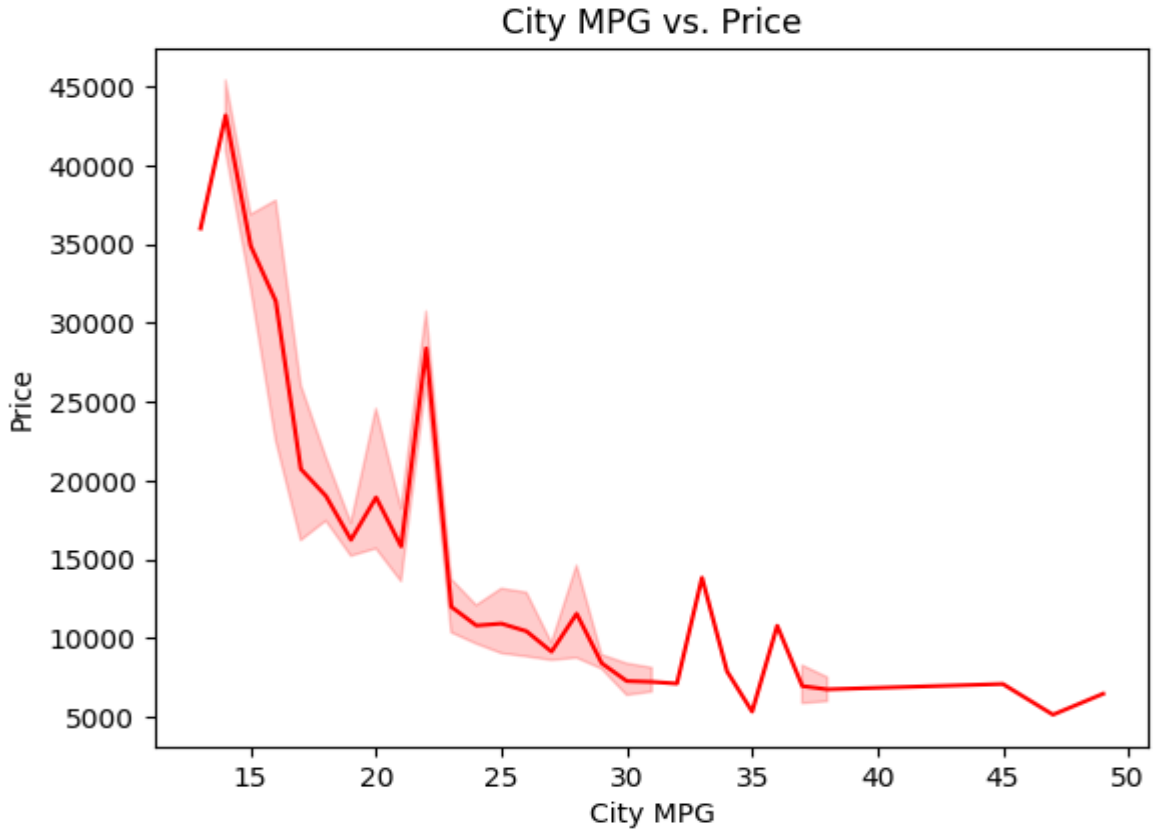




Cars with hatchback and sedan body styles are more fuel efficient than those with hardtop or convertible.

City MPG vs Price (Scatterplot and Line plot)





Car price and fuel efficiency have a strong negative relationship, this means that expensive cars are less fuel efficient than cars that are cheaper.

CONCLUSION

This analysis provides valuable insights into the fuel efficiency of cars in the Dataset and the factors that affect efficiency. The conclusion of the analysis has indicated that Chevrolet is the most fuel-efficient car make in the dataset, followed by Isuzu with jaguar being the least fuel-efficient car in the Dataset. So, Chevrolet and Isuzu produced the most fuel-efficient cars. The is a strong negative relationship between the variables, this means that for every increase in engine size, the is a decline in fuel efficiency. So, the size of a car's engine does impact fuel efficiency. Cars that have big engine sizes are less fuel efficient. cars that use diesel are more fuel efficient than cars that use gas. Cars with hatchback and sedan body styles are more fuel efficient than those with hardtop or convertible. Car price and fuel efficiency have a strong negative relationship, this means that expensive cars are less fuel efficient than cheaper cars.

THIS REPORT WAS WRITTEN BY: XOLANI JEANPAUL MKHALIPHI